3aMU9. The spectral and acoustical impact of vowel changing in choral tuning

Bertrand Delvaux* and David Howard

*Corresponding author’s address: University of York, NAMUR, 5000, NAMUR, Belgium, bertrand.delvaux@gmail.com

Mixed soft/solid models of the Vocal Tract were moulded with a 3D rapid prototyping technique based on MRI data obtained from two male singers during the phonation of five English vowels as in hard, stern, neap, port and food. The replicas are used to assess the interaction of several vocal tracts in different settings: twice the same singer or two different singers, singing on the same vowel or on different vowels, on a consonant or a dissonant interval. The spectral output is analysed and the acoustical output is submitted to a listening test to evaluate the spectral and acoustical grounds for an interval/chord to perceptually sound “in tune”.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

There have been many studies relating to the singing voice over the last few decades ([1], [2], [3], [4]) but little research relates to choir singing, particularly ensemble intonation.

Ternström and Sundberg [5] have studied the intonation of choir, Sacerdote [6] have measured frequency fluctuations and vibrato in choirs and Howard [7] has investigated pitch drift in a cappella (unaccompanied) quartet singing.

In ensemble singing, we believe that the perceived intonation depends on both the vowels sung and the singers combination. We assess the perceived intonation between two experienced male singers (S1 and S2, bass and tenor respectively) using moulded replicas of the Vocal Tract (VT). These are driven by an input sound signal representing the glottal flow. We analyse the sound output through spectrograms and a listening test.

MATERIALS AND METHODS

Moulding the Vocal Tracts

The VT replicas were moulded based on volumetric MRI data collected while two professional male singers were singing 2 English vowels in a supine position [8], by a 3D fast gradient echo sequence. The relaxation time was 4.8 ms and the excitation time was 1.7 ms. Acquisition is isotropic 2mm in a 192x192 matrix. Output is then interpolated to 512x512 using 50% slice overlap giving an effective anisotropic output of 0.75x0.75x1mm. A stack of 80 images is produced in the midsagittal plane in approximately 16s.

The MRI data were then segmented with the open source code ITK-Snap, to rebuild a 3D Vocal Tract, whose .STL file was then printed in the Electronics Department of the University of York, on Objet24 (a 3D printer made by Objet) using the material called “Vero White Plus FullCure 835” and a support material called ”FullCure 705 support”.

The VTs are composed of an insert to plug onto a 952.210 horn driver and lips of actual shape (see Fig. 1). The thickness of the shell of the VT is 2 mm.

FIGURE 1: The VT replicas, with the insert and the actual shaped lips

Pure Data

Pure Data (Pd) is a graphical programming environment for audio, video, and graphical processing. It is used to generate the glottal waveform injected in the VTs, here a sawtooth signal of the form
\[ x(t) = \sum_{k=1}^{\infty} \frac{\sin(2\pi k f_0 t)}{k} \]  

with \( f_0 \) the fundamental frequency, and \( t \) the time. This infinite series is truncated and normalized. The code in pD allows to fix \( f_0 \), the vibrato rate and the vibrato depth. We fixed the value \( f_0 = 132 \) Hz (C3) and chose a vibrato rate and depth of 5.5 Hz and +/- 30% of a semitone respectively, which are common values accepted in the literature. The scale was tuned according to "just intonation" [9] (see Tab. 1).

**TABLE 1:** Scale used in the experiment, with the frequency (in Hz) and ratio to the root note (C3 = 132 Hz)

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>Db</th>
<th>D</th>
<th>Eb</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>Gb</th>
<th>A</th>
<th>A#</th>
<th>B</th>
<th>Bb</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>132</td>
<td>140.8</td>
<td>148.5</td>
<td>158.4</td>
<td>165</td>
<td>176</td>
<td>190.08</td>
<td>198</td>
<td>211.2</td>
<td>220</td>
<td>237.6</td>
<td>247.5</td>
<td>264</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>16</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>36</td>
<td>35</td>
<td>34</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>15</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>26</td>
<td>25</td>
<td>24</td>
<td>19</td>
<td>18</td>
</tr>
</tbody>
</table>

**Acoustic measurements**

The experiment took place in a 6-sided acoustic anechoic chamber. Both VTs were located 20 cm apart, and a measuring mic MBC 550 was placed 20 cm far from each of them, occupying a vertex of a 20 cm edge equilateral triangle. The recorded output was written on a USB type device with a 96 kHz sampling rate on a 16 bits WAV file. The VTs were given an input signal (as described before) at the glottis by a 952.210UK driver unit.

**RESULTS**

To assess the perceived intonation between two VTs, we chose four harmonic intervals. An interval between two tones is harmonic when the ratio between them can be expressed by two small integers \( m : n \) (\( m, n \)-integers where \( m > n \)). For this study, we chose:

1. a minor third, \( m3^{rd} \) (6:5)
2. a major third, \( M3^{rd} \) (5:4)
3. a perfect fifth, \( 5^{th} \) (3:2)
4. a minor seven, \( m7^{th} \) (9:5)

We evaluated the spectral factors of each of these intervals for 1 singer with two different vowels and 1 vowel with two different singers. This gives 4 sets in total:

1. S1 (bass) with i-a
2. S2 (tenor) with i-a
3. i with S1-S2
4. a with S1-S2

Each set was given the 4 intervals, with 2 settings for each: S1 - S2 or S2 - S1. This gave 16 pairs of sound. For each pair, 11 expert listeners were asked "Which sound do you prefer?". They were given the choice among the first setting, the second setting and no preference.
On Fig. 2, we can see the evaluation given by the 11 expert listeners, for the same singer (i.e., S1 or S2) singing the intervals with ‘a’ or ‘i’ as a root (in blue and in red respectively). S1 exhibits a preference for ‘i’ as the root note for the 5th and ‘a’ as the root note for the m7th (as illustrated on Fig. 3a, where coming the right part (‘i’ as the root) is denser in harmonics). S2 shows a clear preference for ‘i’ as the root (as can be seen on Fig. 3b).

On Fig. 4, we can see that for ‘i’, S2 was generally preferred as a root note whereas for ‘a’, S1 was preferred as the root. The former case is illustrated on Fig. 5a, where the spectrogram gets denser around 2.5 - 3.5 kHz when it goes from S1 ro S2 as the root. The latter case is illustrated on the Fig. 5b, with denser harmonics on the part with S1 as a root.

The listening test confirmed what could be observed on the spectrograms. The results can be summarized as follows:

- For S2, ‘i’ was preferred as a root
- For ‘i’, S2 was generally preferred as a root
- For ‘a’, S1 was preferred as the root.
- The m7th is the interval which listeners rated the most unanimously

**DISCUSSION**

From the summary of results above, we could observe that S2-‘i’ and S1-‘a’ are strong combinations, reinforcing the partials of the other singer/vowel when they are taken as a root note. In other words, our bass singer exhibits a denser spectrum when singing the vowel ‘a’ whereas our tenor singer exhibits a denser spectrum when singing the vowel ‘i’. The vowel ‘i’ was found to generally be better as the root of a harmonic dyad. This is a preliminary study. Therefore, further refinements and a more extensive study including more vowels/singers are needed to find patterns in the spectral role of vowels and singers for ensemble intonation.

**REFERENCES**


**FIGURE 2:** S1 and S2 with 'a' (in blue) or 'i' (in red) being the root note of the interval. In yellow, 'no preference'.

(A) S1, $m7^{th}$. Root note from /a/ to /i/

(B) S2, $5^{th}$. Root note from /a/ to /i/

**FIGURE 3:** Spectrograms
**FIGURE 4:** 'i' and 'a' with S1 (in blue) or S2 (in red) being the root note of the interval. In yellow, 'no preference'.

(A) 'i', 7th. Root note from S1 to S2  
(B) 'a', 5th. Root note from S1 to S2

**FIGURE 5:** Spectrograms
FIGURE 6: Ascending scales

(A) S1, root note from /a/ to /i/

(B) S2, root note from /a/ to /i/

FIGURE 7: Ascending scales

(A) 'i', root note from S1 to S2

(B) 'a', root note from S1 to S2